POWER MANAGEMENT FOR HETEROGENEOUS COMPUTING SYSTEMS

BACKGROUND

[0001] Field of the Disclosure

[0002] The present disclosure relates generally to large-scale computing systems and, more particularly, to power management in large-scale computing systems.

[0003] Description of the Related Art

[0004] The costs and technical difficulties of distributing sufficient power among the servers of a data center, along with the corresponding cooling requirements, have given rise to power management systems that seek to maintain a specified power budget or thermal envelope through server consolidation, job migration, and power capping. However, conventional approaches to power management assume a homogeneous system, that is, that servers of the same type or class exhibit the same power dynamics. This assumption, and the power management approaches it encourages, often leads to sub-optimal processing performance for a given power budget.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] The present disclosure may be better understood, and its numerous features and advantages made apparent to those skilled in the art by referencing the accompanying drawings. The use of the same reference symbols in different drawings indicates similar or identical items.

[0006] FIG. **1** is a block diagram of a data center utilizing a heterogeneous-aware power management system in accordance with some embodiments.

[0007] FIG. **2** is a flow diagram illustrating a method for power management in a data center based on per-server power dynamics in accordance with some embodiments.

[0008] FIG. **3** is a flow diagram illustrating a method for implementing various power management configurations in a data center based on per-server power dynamics in accordance with some embodiments.

DETAILED DESCRIPTION OF ONE OR MORE EMBODIMENTS

[0009] Heterogeneity in the power dynamics of individual servers (despite apparent homogeneity in the components or configuration of those servers), racks of servers, and aisles of server racks is the reality of data center operation, even with seemingly homogenous use of similar machines. Various factors contribute to the different power dynamics among individual servers or groups of servers. For one, process variations often result in power performance variations on the order of 10-20% for the same part number or model number. Because of socket compatibility, different part models, with different power dynamics, may be used in seemingly identical machines. Some servers may have different system memory or local storage capacities than others due to failures, which in turn can impact per-watt performance. Further, components tend to exhibit changes in performance as they age, and thus the difference in ages of parts among the servers can contribute to differences in power consumption. Moreover, heterogeneity in server power dynamics may also occur for reasons external to the servers themselves, such as due to individual differences in fan power performance, or the typical differences in cooling efficacy in different regions of the data center.

[0010] FIGS. **1-3** disclose techniques to account for the heterogeneous power dynamics of computing resources in a data center so as to more efficiently implement various power management schemes. In at least one embodiment, each computing resource of a set of computing resources is evaluated to determine a corresponding idle power consumption metric and a peak power consumption metric for the computing resource. The computing resources may be individual servers, or groups of servers, such as a rack of servers, an aisle of racks, a larger subsection of the data-center, and the like. The evaluation of each computing resource can include testing (one more times) of the computing resource, or determination of the power consumption metrics from documentation provided by the supplier of the computing resource. A dynamic power consumption metric, represented by the difference between the peak power consumption metric and the idle power consumption metric, also is determined for each computing resource of the set (one or more times). These power consumption metrics then are used by a power management system to configure the set of computing resources to more efficiently meet a specified power budget constraint, which may represent a maximum power consumption allocated to the set of computing resources, a maximum thermal envelope allocated to the set of computing resources, or a combination thereof. The techniques employed by the power management system may include, for example, server consolidation based on idle power consumption metrics, workload allocation or reallocation based on peak power consumption metrics and dynamic power consumption metrics, power capping based on dynamic power consumption metrics, and the like.

[0011] For ease of illustrations, the power management techniques are described below generally in the example context of computing resources as individual servers. However, the described techniques are not limited to this example, but instead may be employed for groups of servers at various granularities, such as on a per-rack basis, per-aisle basis, per-group-of-aisles basis, and the like.

[0012] FIG. **1** illustrates a data center **100** employing heterogeneous-aware power management schemes in accordance with some embodiments. The data center **100** includes a power management system **102** and one or more sets **104** of two or more computing resources. For the depicted example, the set **104** includes four computing resources: computing resources **111**, **112**, **113**, and **114** (collectively, "computing resources **111-114**"). However, the number of computing resources in a set may be fewer or greater than the four described. In the event that the power management system **102** includes multiple sets **104**, each set **104** may have the same number of computing resources or the sets **104** may have different numbers of computing resources.

[0013] As each computing resource is a server in this example, each computing resource includes one or more motherboards comprising one or more processing units connected to the motherboard via a corresponding socket, system memory, one or more disk drives or other local storage devices, a network interface card, and the like. The power management system **102** is connected to each of the computing resources via one or more data networks **116**, such as an Ethernet network, used for communicating workload information and data between the power management system **102** and the computing resources, as well as among the computing resources. The power management system **102** further is connected to the computing resources via a